

A Survey of Computer Methods in Forensic Document Examination

Sargur N SRIHARI¹ and Graham LEEDHAM²

¹*Center of Excellence for Document Analysis and Recognition (CEDAR),
University at Buffalo, State University of New York, Buffalo, USA. srihari@cedar.buffalo.edu*

²*School of Computer Engineering, Nanyang Technological University,
Singapore. asgheedham@ntu.edu.sg*

Abstract. Forensic document examination is at a cross-roads due to challenges posed to its scientific basis as well as due to the availability of revolutionary computer methods. This paper surveys efforts to establish the scientific basis of forensic document examination --some of which are based on computational theories-- as well as software tools to assist document examiners. They include: computer systems to provide the degree of match of questioned and known documents, systems that narrow-down the search from a repository of documents with known writers and tools that compute features and provide visualization to assist the document examiner.

1. Introduction

The analysis of handwriting from the viewpoint of identification of the writer has a long history perhaps dating to the origins of handwriting itself. Crime involving documents, ranging from fraud and anonymous letters to armed robbery and murder, is dealt with by Questioned, or Forensic Document Examiners. Much of this forensic work involves the comparison of handwriting and handwritten signatures. Numerous techniques have been developed and employed during the past century in the examination of handwriting to establish (i) the identity of the writer of a questioned document from one or a set of known handwritings; (ii) whether the handwriting is forged (the author is not who he claims to be) or (iii) whether the handwriting is disguised (the author is trying to disclaim the writing). There exist numerous text-books describing the general methodology employed by forensic document examiners in various developed countries, e.g., Robertson, 1991; Russell & Bradford, 1992; Harrison, 1981; Hilton, 1993; Huber & Headrick, 1999.

Whether handwriting can be used to identify a person is of great importance to justice and law enforcement systems. Other branches of forensic examination of evidence, such as DNA analysis, analysis of fibers and other material is supported by a wealth of chemical and biological scientific knowledge which has been demonstrated and proven by scientific study. During the last decade numerous challenges have been made regarding presenting expert forensic document examination testimony in the courts (Daubert vs Merrell Dow Pharmaceuticals, 1993, United States vs Starzecypzel, 1995). The challenges are based on whether handwriting evidence has a scientific validation of its individuality. Long-established forensic handwriting examination has only recently faced this question and it has been found that there are inadequate scientific studies for the individuality argument. Such challenges, known in the United States as Daubert challenges, have led to a need for a scientific demonstration of the individuality of handwriting. The problem is that forensic document examination employs many reasonable but scientifically unproven techniques. The acceptability of expert opinion is strongly based on the credibility and standing of the document examiner rather than on the scientific evidence supporting their opinion.

Over the past 30 years there has been a limited amount of research into using computers to enhance and automate the analysis performed by forensic document examiners (Kukuck et al. 1979; Black, 1992; Behnen & Nelson, 1992). Much of the research centered on pattern recognition techniques for extracting static and dynamic features from handwriting and hand written signatures as well as enhancing document images and ESDA lifts (Baier et al., 1987). These were primarily software tools to assist a document examiner rather than establishing any scientific basis for their work.

In recent years forensic document examiners in various countries, and particularly in the USA, have been seeking to strengthen their standing through Accreditation and Certification schemes, formalized training and other means (FSAB, 2003).

More recently, computer scientists have begun to apply various computer-vision and pattern recognition techniques that have been developed during the past 40 years (Plamondon & Srihari, 2000), to the problems of writer identification and the authenticity/individuality of handwriting.

In this paper we review some the key techniques and results that have been published over the past few years in providing scientific support and computer-based tools to assist forensic document examination. This paper focuses on articles published in English, although it is recognized that some key relevant work is published in other languages. A future paper will address this larger body of work.

2. Computational methods to establish individuality

Establishing the scientific basis of forensic document examination has been approached recently in two different ways. The first of these is to establish that the performance of expert forensic document examiners is superior to that of lay people. This has been done by performing controlled tests with human subjects (Kam, et. al., 1994; 1997; 1998 and 2001). Kam et al. concluded that forensic document examiners are significantly better at identifying valid and forged handwriting than the lay person. This indicates that the training and methods used by professional document examiners is indeed a sign of expertise. However, their ability and performance was only gauged on proficiency tests. There is no clear scientific evidence on which they can fully support their decision.

Another approach to establishing a scientific basis for handwriting examination is through the development of a computational theory. A computational theory is an application of computer vision which consists of representations, algorithms and implementations. The computational theory approach has the advantage of repeatability, i.e, the same results are obtained when applied to the same documents, as opposed to expert human document examiners. Based on a computational theory and a particular realization of that theory through specific algorithms and software implementations have led to a relatively high confidence in the individuality of handwriting (Srihari, et. al., 2002).

A computational theory of handwriting examination has been developed and the theory tested with experiments for the task of handwriting identification as well as for handwriting verification (one where the task is to determine whether two documents were written by the same individual or by different individuals) (Srihari, et. al., 2002; Cha, 2001; Cha & Srihari, 2000; Lee et al., 2002). These experiments used both *macro*-features, obtained at the global level from the entire document, and *micro*-features, obtained at the level of individual handwritten characters.

The degree to which handwriting is individual has been explored in the context of handwritten numerals (Srihari, et. al., 2003a), alphabets and words (Zhang & Srihari., 2003b). In these studies the degree to which numerals, alphabets and words are useful in discriminating between individuals are given.

The stability of features used by document examiners for letter-level comparison have also been examined (Sutanto et al., 2003; Leedham et al., 2003a, 2003b) as well as overall writer identification methods (Greening, 1998). The letter-level studies, based on five different lower case letters ('a', 'd', 'f', 't' and 'y'), indicate that there is stability and individuality in some of these features, thus supporting the observations of Kam et al. and the techniques employed by professional forensic document examiners.

3. Software tools to assist document examiners

Several computer software tools to assist document examiners (but not provide scientific support) in their analysis techniques were developed in the 90's. One particularly effective system (known as the FISH system) was developed by the German law enforcement (Hecker, 1993) enabled retrieval of the closest match from a large database of examples handwritings. Current research is further developing this trend in the Wanda Project (Franke et.al., 2003).

A number of tools were researched and developed for separation of the handwriting from the background (paper) in scanned document images to ensure that the detail of the handwritten strokes is retained (Leedham & Sagar, 1994, 1995; Solihin & Leedham, 1997a, 1997b, 1999; Leedham, 1999) as well as perform semi-automatic line, word and letter segmentation of the handwritten script for comparison chart generation (Holcombe et al., 1995, 1996). Tools were also provided to perform global feature extraction and provide visualization for the individual letters/words of the handwriting (Chong et al., 1996a, 1996b; Greening et al., 1996).

Current commercial products available to assist document examiners include the *WriteOn* and *Pikaso* software systems.

A complete system for handwriting examination known as CEDAR-FOX has been recently developed (Srihari, et. al., 2003b). It can be used in several modes of operation: (i) *identification*, where the goal is to identify the writer of a questioned document given a repository of writing exemplars of several known writers, (ii) *verification*, where the goal is to provide a level of confidence as to whether a questioned document and a known document are from the same writer, and (iii) *analysis*, where the system allows the document examiner to perform several image processing operations to analyze the questioned document(s).

Central to both identification and verification is the need for associating a quantitative measure of similarity between two samples. Such a quantitative measure brings in an assurance of repeatability and hence a degree of objectivity. Several methods for comparing strings of binary feature vectors representing handwritten characters have been recently evaluated (Zhang & Srihari, 2003a, 2003b). This has led to the use of a correlation measure of binary string matching within the CEDAR-FOX system.

4. Future research directions and challenges

The research into computer methods in forensic document examination is still at an early stage. The research reviewed is only concerned with techniques and tools for the comparison of writing samples to identify authorship. There is considerable further research required to provide detailed scientific evidence of the nature of handwriting individuality.

Further research is also needed to provide scientific evidence and tools to identify disguised writing, forged handwriting and well as altered or modified writing. In addition, computers could provide impressive assistance in the restoration or decipherment of damaged or partially destroyed documents.

5. References

- Baier, P.E., Teder, W., Hussong (1987). Future Trends in Automatic Document Analysis, *Symposium on Handwriting and Computer Applications*, Montreal, Canada.
- Behnen, A. P., Nelson, L. K. (1992). Additional Applications of Digital Image Processing to Forensic Document Examinations, *Journal of Forensic Science*, JFSCA, Vol 37, No. 3, 797-807
- Black, J. A. (1992). Application of Digital Image Enhancement Software with the Macintosh Computer to Questioned Document Problems, *Journal of Forensic Science*, JFSCA, Vol. 37, No. 3, 783-796
- Cha, S-H., S N Srihari (2000). Writer Identification: Statistical Analysis and dichotomizer, *Proc. SS&SPR 2000 LNCS – Advances in Pattern Recognition*, Springer Verlag, Vol. 1876, 123-132.
- Cha, S-H. (2001). Use of distance measures in handwriting analysis, *PhD dissertation*, State University of New York at Buffalo.
- Chong S.W., Leedham C.G., Solihin Y., Sagar V.K. (1996a). Slant Manipulation and Character Segmentation for Forensic Document Examination, *Proceedings the IEEE Region Ten Conference: Digital Signal Processing Applications*, Perth, Nov 27-29, 933-938.
- Chong S.W., Leedham C.G., Sagar V.K. (1996b). FOX - Forensic Document Examination Tools *Proc 5th International Workshop on Frontiers in Handwriting Recognition*, Essex, UK, 2nd-5th September 1996, 213-220.
- Daubert vs Merrell Dow Pharmaceuticals (1993). US Court Supreme Court Ruling, 509 US 579.
- Franke K., Schomaker L., Vuurpijl L., Giesler St. (2003). FISH-new: a common ground for computer-based forensic writer identification, *Third European Academy of Forensic Science Triennial Meeting*, Istanbul, Turkey.
- FSAB (2003). Forensic Specialties Accreditation Board, <http://thefsab.org/>
- Greening, C. (1998). Automatic writer identification for forensic document analysis, *PhD dissertation*, University of Essex, UK
- Greening C, Sagar V.K., Leedham C.G. (1996). Forensic Document Examination using Automatic Separation of Handwritten Text and Machine Printed Text, *In Handwriting and Drawing Research: Basic and Applied Issues*, M. L. Simner, C.G. Leedham, A. J. W. M. Thomassen (Eds.), IOS Press, 509-520.
- Harrison W.R. (1981). *Suspect Documents, their Scientific Examinations*, Nelson-Hall Inc., Illinois.
- Hecker M.R. (1993). Forensic Information System for Handwriting (FISH), *Technical Document from the Kriminaltechnisches Institut, Bundeskriminalamt*, 6200 Wiesbaden, Germany
- Hilton O. (1993). *Scientific Examination of Questioned Documents*, CRC Press, Florida.
- Holcombe G., Leedham C.G., Sagar V.K. (1995). Image Processing Tools for the Interactive Forensic Examination of Questioned Documents, *Proceedings European Convention on Security and Detection*, Brighton, UK, 16-18 May 1995, 225-228.

- Holcombe G, Leedham C.G., Sagar V.K. (1996). Computer Aided Creation of Handwriting Comparison Charts in the Forensic Examination of Questioned Documents, *In Handwriting and Drawing Research: Basic and Applied Issues*, M. L. Simner, C.G. Leedham, A. J. W. M. Thomassen (Eds.) IOS Press, 493-507.
- Huber R.A., Headrick A.M. (1999). *Handwriting Identification: Facts and Fundamentals*, Boca Roton: CRC Press.
- Kuckuck, W., Rieger, B., Steinke, K. (1979). Automatic Writer Recognition, *Proc. 1979, Carnahan Conf. on Crime Countermeasures*, University of Kentucky, Lexington.
- Kam M, J. Wetstein, R. Conn (1994). Proficiency of professional document examiners in writer identification, *Journal of Forensic Sciences*, JFSCA Vol 39, 5-14.
- Kam M., Fielding G., Conn R. (1997). Writer identification by professional document examiners, *Journal of Forensic Science*, 42(5), 778-786.
- Kam M., Fielding G., Conn R. (1998). Effects of monetary incentives on performance in document examination proficiency tests, *Journal of Forensic Science*, 43(5), 1000-1004.
- Kam M., Gummadidala K, Fiedling G., Conn R. (2001). Signature authentication by forensic document examiners, *Journal of Forensic Science*, 46(4), 884-888.
- Lee, S., S-H. Cha, S. N Srihari (2002). Combining Macro and Micro Features for Writer Identification, *Proceedings of SPIE*, Jan 2002.
- Leedham C. G. (1999). Image Analysis Tools for Authentication and Enhanced Classification of Handwritten Script Using Forensic Techniques, *Final report of research grant RG25/95* from the Singapore Universities Research Fund, May 1999.
- Leedham C.G., Sagar V.K. (1994). Using Forensic Handwriting Analysis Techniques to Enhance Automatic Handwritten Script Recognition and Processing, *Proceedings the IEE European Workshop on Handwriting Analysis and Recognition: a European Perspective*, Brussels.
- Leedham C.G., Sagar V.K. (1995). Image Processing Techniques in the Forensic Analysis of Handwriting, Invited presentation at the *Electronic Imaging and Forensic Science Conference*, London, UK.
- Leedham C.G., Pervouchine V., Tan W.K., Arun J. (2003a). Assessment of the stability and usefulness of some handwriting features used by document examiners to identify authorship, *Proc 11th International Graphonomics Society Conference*, Scottsdale, USA, This volume.
- Leedham C.G., Pervouchine V., Tan W.K., Arun J. (2003b). Automatic quantitative letter-level feature extraction of document examiner features, *Proc 11th International Graphonomics Society Conference*, Scottsdale, USA, This volume.
- Plamondon, R., Srihari, S. N. (2000). On-line and off-line handwriting recognition: A comprehensive survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1): 63-84.
- Robertson E.W. (1991), *Fundamentals of Document Examinations*, Nelson-Hall Inc., Illinois.
- Russell R Bradford, Ralph B Bradford (1992). *Introduction to handwriting examination and identification*, Nelson-Hall Publishers, Chicago.
- Solihin Y., Leedham C.G. (1997a). Noise and Background Removal from Handwriting Images, *Proceedings of IASTED International Conference on Intelligent Information Systems*, ISBN 0-8186-8218-3, IEEE Computer Society, 366-370, The Bahamas.
- Solihin, Y., C.G. Leedham (1997b). Preparing Handwriting Images for Forensic Examination, *Proceedings of IGS'97 - 8th Biennial Conference of the International Graphonomics Society*, Genova, Italy, 24-28 August 1997, 125-126.
- Solihin Y., Leedham C.G. (1999). Integral Ratio: A New Class of Thresholding Techniques for Handwritten Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 8, August 1999, 761-768.
- Srihari, S.N., S-H Cha, H. Arora, S. Lee. (2002). Individuality of Handwriting, *Journal of Forensic Sciences*, 44(4): 856-72.
- Srihari, S.N., C. I. Tomai, B. Zhang, S-J. Lee (2003a). Individuality of Numerals, *Proceedings International Conference on Document Analysis and Recognition (ICDAR)*, Edinburgh, Scotland, August 2003, 1096-1100.
- Srihari, S.N., B. Zhang, C. Tomai, S-J. Lee, Z. Shi, Y.C. Shin (2003b). A system for handwriting matching and recognition, *Symposium on Document Image Understanding Technology (SDIUT 03)*, Greenbelt, MD.
- Sutanto P.J., Leedham C.G., Pervouchine V. (2003). Study of the consistency and some discriminatory features used by document examiners in the analysis of handwritten letter 'a', *Proc. 7th International Conference of Document Analysis and Recognition*, 3-6 August, Edinburgh, Scotland, 1091-1095.
- United States vs Starzecypzel (1995). 880 F Supplement. 1027, 1046 (SDNY).
- Zhang, B., Srihari, S.N. (2003a). Binary vector dissimilarity measures for handwriting identification, T. Kanungo, Smith, E. H. B., Hu, J. and Kantor, P.B., eds., *Document Recognition and Retrieval X*, Bellingham, WA: SPIE, 5010: 28-38.
- Zhang, B., S. N. Srihari (2003b). Analysis of handwriting individuality using word features, *Proceedings International Conference on Document Analysis and Recognition (ICDAR)*, 3-6 August, Edinburgh, Scotland, 1142-1146.